Post-translational regulation of activation-induced cytidine deaminase

Uttiya Basu\(1,2,3\), Andrew Franklin\(1,2,3\) and Frederick W. Alt\(1,2,3,*\)

1Howard Hughes Medical Institute, The Children’s Hospital, Boston, MA 02115, USA
2Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
3Immune Disease Institute, Boston, MA 02115, USA

The assembled immunoglobulin genes in the B cells of mice and humans are altered by distinct processes known as class switch recombination (CSR) and somatic hypermutation, leading to diversification of the antibody repertoire. These two DNA modification processes are initiated by the B cell-specific protein factor activation-induced cytidine deaminase (AID). AID is posttranslationally modified by phosphorylation at multiple sites, although functional significance during CSR has been implicated only for phosphorylation at serine-38 (S38). Although multiple laboratories have demonstrated that AID function is regulated via phosphorylation at S38, the precise biological role of S38 phosphorylation has been a topic of debate. Here, we discuss our interpretation of the significance of AID regulation via phosphorylation and also discuss how this form of AID regulation may have evolved in higher organisms.

Keywords: activation-induced cytidine deaminase; class switch recombination; evolution; phosphorylation; somatic hypermutation

1. INTRODUCTION

The DNA sequences that encode the variable regions of immunoglobulin (Ig) heavy (H) and light (L) chains are assembled from variable (V), diversity (D) and joining (J) DNA segments during early B cell development (Dudley et al. 2005). The products of recombination activating genes 1 and 2 are expressed in a lymphocyte-specific manner and initiate V(D)J recombination by recognizing and introducing DNA double-strand breaks (DSBs) at the recombination signal sequences that flank germ line V, D and J segments. Subsequently, non-homologous end-joining fuses broken V, D and J segments (Dudley et al. 2005). Productive assembly of an IgH variable region exon places a VDJ exon upstream of the CH constant region exons, allowing generation of µ heavy chain (HC) mRNA and µHC proteins. The µHC pairs with IgL chains produced from a functionally rearranged IgL locus to form IgM, leading to generation of surface IgM+B cells. IgM+B cells migrate to secondary lymphoid organs where, upon encounter with cognate antigens, they are activated to undergo class switch recombination (CSR) and somatic hypermutation (SHM). CSR is a recombination/deletion process that replaces Cµ with a different set of IgH constant region exons (Chaudhuri et al. 2007) and SHM introduces point mutations into variable region exons (Di Noia & Neuberger 2007). Both CSR and SHM in mice and humans are initiated by activation-induced (cytidine) deaminase (AID; Muramatsu et al. 2000).

2. CLASS SWITCH RECOMBINATION

The Ig HC constant (C\(\mu\)) region determines biological effector functions of Ig molecules. In mice there are eight sets of C\(\mu\) exons (C\(\mu\) genes) as organized as 5'-VDJ-C\(\mu\)-C\(\delta\)-C\(\gamma\)3-C\(\gamma\)1-C\(\gamma\)2b-C\(\gamma\)2a-C\(\epsilon\)-C\(\alpha\)-3'. Mice produce five IgH classes, IgM, IgD, IgG, IgE and IgA, encoded by \(\mu\), \(\delta\), \(\gamma\) and \(\epsilon\) C\(\mu\) genes, respectively. Each C\(\mu\) region is specialized in determining how an Ig molecule eliminates antigen. CSR allows the assembled VDJ exon, initially expressed with the Cµ exons, to be expressed with one of the sets of downstream C\(\mu\) exons (Chaudhuri et al. 2007), and hence enables the production of different IgH classes. CSR occurs within large, repetitive switch (S) regions located 5' of each set of C\(\mu\) exons. CSR is a deletional recombination event that occurs via the introduction of DSBs into the donor S\(\mu\) and a downstream acceptor S region, followed by joining of broken donor and acceptor S regions to each other (Honjo et al. 2002).

Mammalian S regions are 1–10 kb in length, unusually G-rich on the template (transcribed) strand, and primarily composed of tandem repetitive sequences within which certain motifs predominate (Chaudhuri & Alt 2004). S\(\mu\) is exceptionally repetitive and enriched in GAGCT motifs, with the AGCT palindrome representing a canonical RGYW motif (where R=purine, Y=pyrimidine and W=A or T). The RGYW motif and the reciprocal WRKY motif are
SHM hotspots (see later). Sγ1 is the largest S region and, like other Sγ sequences, carries multiple 49 bp repeats that are rich in RGYW motifs (Chaudhuri & Alt 2004). In some species, S regions are quite divergent; Xenopus S regions are AT-rich (Du Pasquier et al. 2000). The vast majority of CSR junctions occur within or, sometimes, just beyond S regions (Dunnick et al. 1993). In normal B cells, deletion of the Sμ repeats greatly reduced CSR (Luby et al. 2001; Khamlichi et al. 2004) and deletion of the 10 kb Sγ1 abrogated CSR to Cy1 (Shinkura et al. 2003). Thus, S regions appear to be the main CSR targets.

The different Cγ exons are organized into units in which germ line transcription initiates from a promoter lying upstream of individual S regions (Chaudhuri et al. 2007). Germ line transcription is a prerequisite for CSR (Manis et al. 2002). In vivo, CSR is stimulated by T cell-dependent and T cell-independent antigens, which can be mimicked in vitro by activating B cells with anti-CD40 or bacterial lipopolysaccharide in the presence of cytokines such as interleukin-4 (IL-4; Chaudhuri & Alt 2004). Different activators and cytokine combinations direct CSR to particular S regions by stimulating germ line transcription at the promoter upstream of particular target S regions (Manis et al. 2002; Chaudhuri et al. 2007).

3. SOMATIC HYPERMUTATION

SHM introduces point mutations, and sometimes small insertions or deletions, at a very high rate (estimated to be approx. $10^{-3}$–$10^{-4}$/base pair/generation) into the variable region exons of IgH and IgL genes, allowing selection of B cells with increased affinity for antigen (Li et al. 2004). SHM occurs preferentially around RGYW and WRCY motifs (Rogozin & Kolchanov 1995), which are particularly abundant in the DNA sequences that encode the variable region complementarity-determining regions (Wagner et al. 1995), the site of the majority of the antigen contact residues. SHM requires transcription (Fukita et al. 1998), with mutations occurring from 100–200 bp from the promoter and extending 1.5–2 kb downstream, generally sparing Cγ exons (Longerich et al. 2006). In transgenic SHM substrate studies, variable region exons or their promoters can be replaced with other sequences or non-Ig promoters with varying degrees of SHM (Storb et al. 2001; Odegard & Schatz 2006), indicating variable region exons are not unique with respect to SHM targeting. SHM can also occur to varying degrees at non-Ig genes, although at much lower levels (Odegard & Schatz 2006; Liu et al. 2008). Precise sequence requirements and cis-acting elements that target SHM in vivo remain to be defined (Odegard & Schatz 2006).

4. ROLE OF ACTIVATION-INDUCED CYTIDINE DEAMINASE RECOMBINATION AND SOMATIC HYPERMUTATION

AID is a 24 kDa protein that is expressed predominantly in activated B cells and is absolutely required for CSR and SHM in mice and humans (Muramatsu et al. 2000; Revy et al. 2000). While the role of AID in CSR and SHM in vivo has been debated, it has become clear that AID functions in these processes by the deamination of cytosine residues in S regions and variable region exons to uracil (Chaudhuri et al. 2007). Genetic studies indicate that AID-generated uracil residues in S regions and variable region exons are processed by base excision repair (BER) and mismatch repair (MMR) to generate DSBs and point mutations, respectively (Di Noia & Neuberger 2007). Other evidence in support of the DNA deamination model include the ability of AID to mutate DNA in bacteria (Petersen-Mahrt et al. 2002); the biochemical demonstration that purified B cell AID (Chaudhuri et al. 2003) and recombinant AID (Branstetter et al. 2003; Dickerson et al. 2003; Ramiro et al. 2003; Sohail et al. 2003) have robust single-stranded DNA (ssDNA)-specific cytidine deaminase activity; and the demonstration that AID binds, in a CSR-dependent fashion, to S regions in activated B cells (Nambu et al. 2003; Chaudhuri et al. 2004). In addition, AID preferentially deaminates cytosine residues within WRC sequences in vivo (Pham et al. 2003). While AID deaminates ssDNA substrates in vitro, it does not deaminate double-stranded DNA (dsDNA) substrates (Chaudhuri et al. 2003). Several mechanisms have been implicated in AID access to duplex V(D)J exons and S regions in vivo as discussed in the following paragraphs.

Owing to their high G/C content and G richness on the non-template strand, mammalian S regions generate ssDNA within R-loops transcribed in the physiological direction (Tian & Alt 2000; Yu et al. 2003), suggesting that transcription might allow AID to access ssDNA substrates by this mechanism, a notion supported by finding that optimal S region function depends on their transcriptional orientation in vivo (Shinkura et al. 2003). Moreover, in vitro purified AID deaminates the non-template strand of T7 RNA polymerase-transcribed dsDNA sequences that form R-loops (e.g. S regions in sense orientation) but not transcribed dsDNA sequences that do not form R-loops (e.g. V(D)J exons; Chaudhuri et al. 2003). Thus, R-loop forming ability may have evolved in mammalian S regions to enhance AID access during CSR.

In vitro assays to identify factors that allow purified B-cell AID to deaminate transcribed SHM substrates (do not form R-loops) revealed that replication protein A (RPA), a trimeric ssDNA-binding protein involved in replication and repair, serves as a cofactor to promote efficient AID deamination of non-R-loop forming dsDNA ‘SHM substrates’ that contain AGCT motifs; (a canonical RGYW motif; Chaudhuri et al. 2004). The 32 kDa subunit of RPA interacts with purified B-cell AID, and the AID/RPA complex binds to transcribed WRCY/RYGW-containing DNA in vitro, hence leading to deamination within or near SHM motifs (Chaudhuri et al. 2004). The AID/RPA complex may bind to and stabilize ssDNA within transcription bubbles to allow AID to access ssDNA substrates (Chaudhuri et al. 2004).

Both AID and Ig gene SHM are present in bony fish. However, CSR first occurs evolutionarily in amphi- bians, suggesting that CSR evolved after SHM (Stawnezer & Amemiya 2004). Notably, Xenopus Sμ (XSμ) is A/T rich and does not form R-loops when transcribed in vitro (Zarrin et al. 2004). Yet, XSμ can...
replace mouse Sγ1 to promote substantial CSR in mouse B cells (Zarrin et al. 2004). CSR junctions within XSβ in mice occurred in a region of densely packed AGCT sequences in the 5′ portion of XSβ; when this sequence was inverted in vivo, CSR junctions tracked with the AGCT motifs (Zarrin et al. 2004). Moreover, the AGCT repeat region was the predominant site of deamination by AID/RPA in vitro, regardless of orientation (Zarrin et al. 2004). Overall, these findings support the notion that the CSR activities of AID evolved from SHM functions (Barreto et al. 2005; Wakae et al. 2006), the latter of which may, at least in part, employ the ability of the AID/RPA complex or related mechanisms to access transcribed sequences rich in SHM motifs (Chaudhuri et al. 2004; Zarrin et al. 2004). Various findings, including the fact that S regions are rich in WRCY/RGYW motifs (particularly AGCT motifs), suggested the possibility that the AID/RPA mode of targeting also may play a role in mammalian CSR (Chaudhuri et al. 2004).

5. POST-TRANSLATIONAL MODIFICATION OF ACTIVATION-INDUCED CYTIDINE DEAMINASE

The question of how AID preferentially targets certain regions within Ig loci and not many other genomic sequences remains unanswered, although transcriptional regulation and target motif density may be contributing factors (Longerich et al. 2006). Other intriguing questions are why AID elicits point mutations in V(D)J exons and DSBs (in addition to point mutations) in S regions, as well as how B cells can differentially activate CSR versus SHM within narrow regions of the IgH locus. The density of target motifs is one simple explanation for DSB versus SHM activity, but this remains to be proven experimentally (Chaudhuri et al. 2007). Differential transcription of target loci directs AID to different S regions during CSR (Chaudhuri et al. 2007) and might similarly help target AID to variable region exons during SHM versus adjacent S regions during CSR. Another potential AID regulatory mechanism might involve the channelling of deaminated cytidines into normal versus abnormal BER or MMR processing (Di Noia & Neuberger 2007). In addition, assays of ectopically expressed mutant forms of AID suggested that particular AID domains might interact with specific cofactors or repair factors to differentially effect SHM or CSR (Barreto et al. 2003; Ta et al. 2003; Shinkura et al. 2004; Imai et al. 2005), potentially in the context of different AID post-transcriptional modifications.

Post-translational modifications influence AID activity in vitro. The ability of AID isolated from activated murine B cells to interact with RPA and function in the transcription-dependent dsDNA deamination assay is dependent on phosphorylation at Serine-38 (S38; Basu et al. 2005). AID S38 lies within a cAMP-dependent protein kinase A (PKA) consensus motif and, correspondingly, AID can be phosphorylated in vitro and in vivo at S38 by PKA (Basu et al. 2005; Pasqualucci et al. 2006). PKA-phosphorylated AID gains the ability to bind RPA and mediate dsDNA deamination of transcribed SHM substrates (Basu et al. 2005). Mutation of AID S38 to alanine (AID S38A) spares ssDNA deamination activity, but markedly decreases PKA phosphorylation, and RPA interaction and function in transcription-dependent dsDNA deamination assays (Basu et al. 2005). Retrovirally expressed AID S38A had reduced CSR activity (15–30% of WT activity) in activated B cells (Basu et al. 2005; McBride et al. 2006; Pasqualucci et al. 2006). Likewise, PKA inhibition decreased CSR in vivo, while deletion of the PKA-negative regulatory subunit increased CSR in vivo (Pasqualucci et al. 2006). Thus, current studies support the proposal that AID phosphorylation at S38 is one mechanism for augmenting the ability of AID to initiate CSR, supporting R-loop-independent means for accessing transcribed S regions (Basu et al. 2005). In addition, retrovirally expressed AID S38A had reduced capacity to mediate SHM in both mouse and chicken DT40 cells (McBride et al. 2006; Chatterji et al. 2007). Notably, however, zebrafish AID (zAID) lacks an S38-equivalent PKA phosphorylation site (figure 1); yet, zAID can rescue CSR in AID-deficient mouse B cells (Barreto et al. 2005; Wakae et al. 2006).Superficially, the latter finding may be viewed as a challenge to the notion that S38 phosphorylation has a physiological function in the regulation of CSR activity (discussed below; Shinkura et al. 2007; Pham et al. 2008).

6. PHYSIOLOGICAL RELEVANCE OF PHOSPHORYLATION

The potential role of AID phosphorylation at S38 has been controversial. Thus, work from the Honjo laboratory questioned whether AID was significantly phosphorylated (potentially due to insensitivity of the assay used) and further concluded, in contrast to our findings and those of others (Basu et al. 2005; McBride et al. 2006; Pasqualucci et al. 2006), that an AID S38A retroviral construct supported CSR at nearly WT levels (Shinkura et al. 2007). We addressed this apparent controversy by comparing CSR activity and expression of AID WT with the corresponding AID S38A constructs generated in our laboratory as well as with AID WT and AID S38A constructs from Shinkura and Honjo (Basu et al. 2007). We found that the Shinkura/Honjo AID WT construct had similar CSR activity to our AID WT construct, but that the Shinkura/Honjo AID S38A

**Figure 1.** AID amino acid multiple alignment for various species corresponding to the region around mouse AID position 38. AID Serine-38 (s), along with the protein kinase A consensus motif (RRXS), is present in all organisms that undergo SHM as well as CSR, whereas the aspartate D44 residue (d) in teleost zebrafish is present only in those organisms that undergo SHM. Homo sapiens (human), Pan troglodytes (chimpanzee), Mus musculus (mouse), Xenopus laevis (frog), Gallus gallus (chicken), Danio rerio (zebrafish), Takifugu rubripes (fugu), Ictalurus punctatus (catfish).

**Phil. Trans. R. Soc. B** (2009)
Figure 2. Schematic representation of the modes by which AID activity is controlled. AID transcription is tightly controlled in B cells by various transcription factors and pathways. In addition, AID mRNA levels in B cells are negatively regulated by the micro RNA miR-155. Phosphorylation of AID at S38 and various other residues has been reported to control its DNA deamination activity. Nuclear levels of AID are regulated by ubiquitination followed by proteosomal degradation. The latter two examples describe the post-translational control of AID via modification(s). Although AID functions in the nucleus, most cellular AID is sequestered to the cytoplasm. AID harbours nuclear export and nuclear localization signals that have now been characterized.

mutant, our AID\textsuperscript{S38A} mutant, had substantially reduced capacity to support CSR (Basu et al. 2007). While the Shinkura/Honjo AID\textsuperscript{S38A} construct did support a somewhat higher level of CSR (35% of WT) than ours (15% of WT), we argued that the difference probably reflected differences in expression levels of the two constructs (Basu et al. 2007). In this regard, the AID\textsuperscript{S38A} mutant retains significant residual CSR activity (Basu et al. 2005), which most likely reflects AID access to S regions via mechanisms that do not depend on S38 phosphorylation (e.g. R-loops; Chaudhuri et al. 2003; Basu et al. 2005, 2007). Thus, assaying expression of such hypomorphic AID mutants under conditions in which AID\textsuperscript{WT} activity is saturated might lead to significant differences in the measurement of hypomorph activity in the context of differences in expression levels and, if over-expressed enough, apparent levels might approach those of WT (Basu et al. 2007). Recently, we have generated mouse B cells that harbour the homozygous S38A mutation in their genomic DNA and, based on preliminary analyses, found that they also had substantially impaired ability to carry out IgH CSR (H.-L. Cheng, B. Vuong, F. W. Alt & J. Chaudhuri 2008, unpublished data).

Another apparently contradictory finding was made by Goodman and colleagues, who recently examined the ability of baculovirus-expressed AID to deaminate an in vitro-transcribed dsDNA substrate and concluded that neither phosphorylation nor RPA was required for AID activity (Pham et al. 2008). Notably, however, inspection of the substrate employed in these studies revealed 45 per cent G residues on the non-template strand, which strongly suggests the potential for the substrate to form R-loops upon transcription. Therefore, this observation may well be in accord with our prior studies in which we had reported that purified AID from mammalian cells could robustly deaminate R-loop containing substrates in the absence of phosphorylation or RPA (Chaudhuri et al. 2003, 2004).

7. EVOLUTIONARY DIFFERENCES IN REGULATION OF ACTIVATION-INDUCED CYTIDINE DEAMINASE VIA PHOSPHORYLATION

Zebrafish and other bony fish possess AID and undergo SHM but do not undergo CSR. Yet zAID can replace function in AID-deficient mouse B cells to catalyse substantial CSR (Barreto et al. 2005; Wakae et al. 2006). Thus, it seems likely that lack of CSR in zebrafish reflects lack of different C\textsubscript{H} genes and S regions (Chaudhuri et al. 2007). Based on the ability of zAID to catalyse CSR in mouse cells, the model that AID activity might be modulated in vivo via a phosphorylation and RPA-dependent mechanism has been called into question by the Honjo and Goodman laboratories based on the observation that zAID lacks a PKA consensus site equivalent to S38 (Shinkura et al. 2007; Pham et al. 2008). On the other hand, we had noted that zAID has an aspartate residue at position 44 (D44) that might serve as a ‘constitutive’ mimic of phosphorylation at S38 of mouse AID (Basu et al. 2005). Notably, all species known to undergo CSR and SHM (humans, mice, amphibians) have a conserved PKA phosphorylation site centred at S38 (or its equivalent; figure 1). Strikingly, bony fish, which are not known to undergo CSR, lack a consensus PKA site centred on an S38 equivalent residue but instead have the conserved D44 residue or its equivalent (Barreto et al. 2005; Wakae et al. 2006; figure 1).

Recently, we have directly addressed the hypothesis that zAID D44 serves as a potential mimetic of phosphorylated S38 in mouse (Basu et al. 2008). In this regard, we have demonstrated that zAID constitutively associates with RPA in the absence of PKA phosphorylation and that association depends on integrity of zAID D44. Moreover, zAID is constitutively active in an in vitro transcription-dependent dsDNA deamination assay, as might be predicted by constitutive RPA association. Likewise, integrity of zAID D44 was required for full zAID activity in a retroviral complementation assay for CSR in AID-deficient mouse B cells. Remarkably, reconstitution of a PKA site into zAID in which the D44 residue was inactivated via an alanine mutation, generated a mutant AID that became phosphorylatable by PKA, showed PKA-dependent RPA-association plus PKA-dependent activity in the transcription-dependent dsDNA deamination assay and which had in vivo CSR activity that approached that of the WT zAID. The proposed functions of RPA in modulating AID activity have been difficult to test in vivo given the critical role of RPA in basic cellular processes such as DNA replication. However, our ability to link an AID second-site mutation to simultaneous gain of PKA phosphorylation-dependent RPA binding, PKA-dependent in vitro dsDNA deamination activity and augmentation of in vivo CSR activity strongly supports the linkage of these three phenomena. Finally, we have used the information from zAID to generate mouse AID mutants that function as a mimetic of AID phosphorylated at S38 (Basu et al. 2008).

8. SIGNIFICANCE AND FUTURE DIRECTIONS

Modest increases in AID levels can significantly affect CSR (Ramiro et al. 2006) and, moreover, can have

Phil. Trans. R. Soc. B (2009)
even greater effect on AID mutator activities, such as promoting translocations (Dorsett et al. 2008; Teng et al. 2008), potentially leading to evolutionary fixation of post-translational control mechanisms to more tightly regulate AID activity (Dorsett et al. 2008; Teng et al. 2008). The regulation of AID via phosphorylation at S38 may add to a growing list of mechanisms that regulate AID that include its transcriptional control, post-transcriptional control by miRNA155 (Dorsett et al. 2008; Teng et al. 2008) and its regulation via nuclear transport and proteosomal degradation (McBride et al. 2004; Aoufouchi et al. 2008; figure 2). In this context, it is not clear whether AID regulation via S38 phosphorylation in higher vertebrates that undergo CSR evolved from a ‘constitutively’ active AID, via the D44 equivalent in bony fish that do not undergo CSR, or if these two forms of AID represent different evolutionary paths from a common ancestral AID species. However, the notion that regulation of AID activity via S38 phosphorylation evolved in concert with CSR, a DSB-generating and therefore dangerous process, is intriguing.

It is notable that the zAID<sup>D44A</sup> mutant, which does not bind RPA, catalyses significant CSR (approx. 50% of WT levels) when expressed in mouse B cells and can catalyse substantial gene conversion of variable region exons in chicken DT40 cells (Chatterji et al. 2007). While the residual in vivo CSR activity of zAID<sup>D44A</sup> and mAID<sup>S38A</sup> may reflect RPA-independent access to ssDNA in S regions in the context of R-loops (Longerich et al. 2006; Basu et al. 2007; Chaudhuri et al. 2007), access to variable region exons, and potentially S regions as well, may also employ additional, potentially related mechanisms. In this context, recent studies using a yeast model suggest that suboptimal co-transcriptional messenger ribonucleoprotein assembly in particular DNA regions may promote AID targeting by contributing to R-loop formation (Gomez-Gonzalez & Aguilera 2007) and similar mechanisms have been proposed for mammalian cells (Li & Manley 2006). An additional mechanism by which ssDNA may be generated at AID target substrates is through formation of stable stem-loop structures in negatively supercoiled DNA behind a translocating RNA polymerase II complex (Wright et al. 2004; Shen et al. 2005).

The authors would like to thank the members of the Alt laboratory for helpful discussions about many of the topics covered in this manuscript. U.B. is a fellow of the Irvington Laboratory for helpful discussions about many of the topics covered in this manuscript. U.B. is an investigator of the Howard Hughes Medical Institute.

**REFERENCES**


Phylog. R. Soc. B (2009)


